USDA

**United States
Department of
Agriculture**

Service Center
Modernization Initiative
(SCMI)

Guide for Describing and Managing Data for Agency Business Systems

Prepared by
Data Management Team #1: Data Administration, Standards, and Policy

Abstract: This document provides guidance to data owners and software developers in describing and managing the data used in agency business systems. Included is a process for establishing, documenting, and managing data; along with the roles and responsibility essential for building and maintaining a valuable information system. Specific management actions are needed at each stage in the System Life Cycle. The document presents a high-level project system lifecycle, with the data management deliverables detailed for each life cycle phase.

Published and distributed on February 27, 2001

Document number: SCMI Gd 006-02

Introduction

The Service Center Initiative (SCI) Data Management Team developed the July 2000 draft of the *Data Management Process* to guide developers in using common processes and procedures for managing data and metadata during the development and implementation of their application.

The individuals who contributed to the development of this guide are:		
David Anderson, (NRCS) Service Center Data Team Leader		
Dave Butler, (NRCS) Leader		
Tom Richards, (FSA) Alt. Leader		
John Bussone (FSA)	Julie Degner (Soza)	Steve Springer (FSA)
Chuck Calcaterra (RD)	Mindy Gehrt (FSA)	Lee Stieger (FSA)
Ruth Davila-Johnston (FSA)	Lynette Harris (NRCS)	Rich Verhoff (RD)

Figure 1 – Working group list

Table of contents

DATA MANAGEMENT

1.   Overview

The objective of any automated information system is to collect, process, and utilize data. Management of this data thereby becomes a cornerstone of all information systems. The process of managing data begins at the inception of a system development project, and continues throughout the life of the data. Data often outlives the system that created it.   Data is an asset usually of considerable value, both monetarily and for its ability to enhance and support agency functions.

1.1 Functions and Responsibilities for Data

This document outlines a basic framework for the duties, responsibilities, deliverables, and procedures needed to deliver an effective ongoing program for managing data. It is useful at this point to explain the difference between Data Management and Data Administration, since the terms are often used interchangeably, but for our purposes are used to denote different functions.

Data Management is the <u>managerial function</u> of taking responsibility for data and the processes that support it. It focuses the strategic planning and data methodologies for meeting program delivery goals. In particular, Data Management aims at managing data as an asset, particularly as a corporate asset. Data managers typically look across applications and business areas to manage the whole architecture of data resources for the enterprise.

Data Administration encompasses the day-to-day <u>technical functions</u> that support ongoing business operations. Each application must implement a data administration process to support system development and ongoing system operation. It includes the collecting, defining, certifying, organizing, protecting, and delivery of both data and metadata (data about data).

The distinction is made because the responsibilities and duties of each function may be delegated and executed differently.   A business area initiates, develops, funds, and implements information systems to support its mission, and ultimately has responsibility for the data assets created by those systems. The business area may hold some data administration functions closely, while others are delegated. Partnerships are developed with IT organizations to provide technical and systems support.

When a business area relies on an IT organization for system development and for data administration support, it is essential for the IT project manager to take the initiative and responsibility to:

- establish the data context for the system, i.e. show how a system and its data fit into the enterprise,
- advise and educate business areas on the procedures to define and manage their data,
- ensure, through the design of software and procedures, that the validity of data is always maintained,
- ensure that a concise and unambiguous definition is created for each piece of data
- take reasonable steps to protect data from being used in ways that invalidate its meaning,
- ensure that the system integrates with the rest of the system enterprise, and
- ensure that systems are designed to promote sharing and reuse of data among systems, business disciplines, and by outside customers.

A number of groups and people may ultimately be involved in the data administration process, including the executive sponsor, data steward, data manager, data administrator, database administrator, project manager, data modeler, programmer, and system user.

Together, the Business Area and the Application Developers should develop a robust plan for the use, management, and protection of the data produced by the system. It is important for Application Developers to budget the time and resources to produce and maintain data documentation deliverables. The workload to produce these deliverables has in the past often not been calculated into project budgets.

1.2 Management Goals for Data

Management Goals

1. Wider distribution of data in ad hoc quantities geared to specific customers. Customers include USDA offices/staff, outside partner agencies, organizations, and individuals.

2. Improved processes for transferring knowledge from universities and research institutes to Service Center offices.

3. More layers, types, and sources of data are readily available, and are catalogued and searchable by subject.

4. Aggregation of different types of data from NRCS offices in order to explore trends. More emphasis on quicker management response to trends. More emphasis on management accountability.

5. Pictorial and spatial emphasis in displaying data.

6. Emphasis on distilling and aggregating data for decision-making.

7. Creation of data to fit a specific need, discarding it when done. Less emphasis on keeping data just for the sake of keeping it. Not trying to integrate everything. Integration needs a business reason.

8. Selection of tabular data by selecting a spatial area of reference, and depicting tabular data as features (points, lines, and polygons) on a map.

9. Increased emphasis on data warehousing technology to organize transactional data for easier retrieval and analysis.

Current Building Blocks for Implementing these Goals

1. A hardware and software environment for prototyping and initial implementations.

2. A gradually-improving telecommunication network for the movement of data, and for wide access to data stores on the web.

3. Identification of specific data elements on which to build permanent and ad hoc inter-project linkages.

4. Establishment of standard tables for highly- or repetitively-used data

5. Standards for data naming, and a required minimum amount of descriptive information about each data element to promote wider usage of data.

6. A searchable repository of information about data elements.

7. Increased knowledge of data structures and organization for warehousing of data.

8. Improved report preparation and distribution methods.

9. Client tools for data mining and exploration.

10. A foundation set of tables and databases on which to build and link information systems.

11. Greater emphasis on planning and managing of data resources.

2   Data Management Process

Management of data is integral to every phase of an information system's lifecycle. Each of the following steps is necessary and requires a certain level of documentation. The documents listed below can be found on the Data Management website http://www.fsa.usda.gov/scdm.

Process Steps	Accomplishments	Related Documents
Develop the concept for a new information system	Establish a business concept, what is expected from an information system, and, in general terms, what are the inputs and outputs.	"Interim Policy for Data Management"
Establish essential roles	Appoint an Executive Sponsor, Data Steward, Project Manager, and Data Team Liaison	See "Roles in Managing Data" in this document
Develop the Initial Project Proposal	Develop a thumbnail sketch of the new system and share it with other business areas and developers. Establish the new system's role and place in the overall information architecture.	"Data Management Deliverables for Application Development Projects" See Appendix B in this document.
Follow the project development lifecycle	Data management activities are built around a generic project lifecycle.	"Data Management Deliverables for Application Development Projects" See Appendix B in this document.
Search for existing sources for data	Don't create data if it already exists and can be shared or re-used.	See Data Management Web Page at http://www.fsa.usda.gov/scdm for data repository search engine.
Identify and define new data elements	Build up a set of data elements to support the mission of the system. If you can't define a data element, it won't be of much use.	"Software Developer's Guide to Standard Data Elements and Common Lookup Tables" "Standard for Service Center Data Naming" "Guide for Identifying and Establishing Standard Data Elements"

Process Steps	Accomplishments	Related Documents
Identify and define new data elements, cont'd		See Data Management Web Page at http://www.fsa.usda.gov/scdm for data repository search engine.
Define Geospatial Data Requirements	Identify spatial themes/layers that will be needed for the system, and how the spatial features will link to tabular datasets.	"Geospatial Data Management Requirements Document" "Standard for Geospatial Dataset File Naming" "A Standard for Geo-referenced Conservation Planning Data" "Standard for Geospatial Data"
Develop the Data Model	Construct logical and physical models for all data used by the system, including data in files and other formats.	"Draft Service Center Data Modeling Standard"
Complete the metadata	Provide at least the minimum metadata for each data element, and each grouping of data.	"Standard for Service Center Tabular Metadata" "Service Center Geospatial Data Set Metadata Standard"
Establish change control	Place the data model and all metadata under change control. Distribute change notices to the system's developers, and to other systems that will link to this new system.	"Draft Policy for Service Center Initiative Configuration Management" "Service Center Agencies Policies for Hardware Configuration, Application Submissions, and Change Control Policy"
Complete the data management deliverables	A minimum set of documentation is essential for future management and maintenance.	"Data Management Deliverables for Application Development Projects"
Establish data stewards at all levels	Prepare for system deployment by appointing & training local data stewards.	"Data Stewardship in the Project Lifecycle"
Upload metadata to the central repository	Upload the system metadata to the central repository.	"Guide for Data Administration of Repository Metadata"
Establish new standard data elements and common tables	Identify particular data elements and tables in the repository that can be shared with other projects.	"Guide for Identifying and Establishing Standard Data Elements"

Process Steps	Accomplishments	Related Documents
Implement a data security plan	Protect sensitive data, and implement the process for backup, recovery, archiving, and transfer to National Archives.	
Deploy the system	Begin gathering and processing real data.	
Document the system updates and fixes	Maintain Change Control. Update system documentation, data models, and the central repository to reflect system changes and updates after deployment.	

3.0   Roles in Managing Data

3.1 Executive Sponsor

The Executive Sponsor is a business-area manager who is accountable for the collection, management, and use of data assets. The Executive Sponsor is the person who determines that data, and the software to collect and manage the data, are necessary to fulfill the business-area mission.

By determining that data is a necessary commodity for achieving the mission, the sponsor assumes overall responsibility for:

     a) coordination of funding for data collection, storage, maintenance, and for software application development, support and maintenance;
     b) establishing standards and policies for the acquisition and certification of data;
     c) establishing the business definition of the data; and
     d) ensuring protection of the physical data assets.

In some cases this may be a shared responsibility between several business-area managers from different agencies or different parts of the same agency.

The Executive Sponsor typically appoints a Data Steward to handle the day-to-day coordination of data management responsibilities, and "contracts" for a project manager to oversee the development and maintenance of software applications.

3.2 Data Steward

The Data Steward is a business-area expert who is assigned responsibility for the content of the database. The Data Steward:

     a) act as the designated authority and point of contact for all business-area decisions concerning the data;
     b) obtains the needs/requirements from the users;
     c) establishes and maintains business rules and consistent definitions for data elements;
     d) develops data quality and certification standards;
     e) determines if data, which the agency plans to collect, has already been collected by the agency, or whether cooperative efforts are possible to obtain the data from other existing sources.
     f) ensures that metadata is collected; and
     g) ensures the validity, accuracy and completeness of the physical data and supporting metadata; certify that data meets quality standards; and certify that data is ready for release for internal and/or public use.

The Data Steward has the ultimate responsibility for: the definition of a data element, the business rules for creating the data, and the proper distribution and usage of the data. The Data Steward must therefore take an active role in the system creation and ongoing management. When the information system is operational, the Steward must actively monitor the creation and validating of the data.

Individual Service Centers are supported by a multitude of Data Stewards, each of whom is responsible for a piece of the total business going on within the Center. Data Stewards are assigned by, and are primarily responsible to, the business-discipline manager within each agency. See Appendix D.

3.3 System Project Manager

The System Project Manager is in charge of system development, implementation, and software maintenance. The Project Manager typically:

a) develops and maintains the system project plan.
b) provides the planning, coordination, staffing and direction to create information systems, and related databases and data stores, to support agency functions and missions.
c) coordinates with the Executive Sponsor and Data Steward to develop information system requirements and design.
d) meets with the data team to develop a project data management plan.
e) ensures that the application is integrated into the enterprise model and data architecture.
f) follows the accepted systems development lifecycle, configuration management, and data management standards and procedures.
g) arranges for the services of a data administrator and a database administrator, as appropriate.
h) works with the Business-Area Data Steward, the Data Administrator, and the Database Administrator during the development process to establish the content, access rules, hardware and commercial software requirements for the information system.
i) incorporates standards for ensuring data integrity into the design of the application systems.
j) shares responsibility with the database administrator and the data steward for the ongoing operation and maintenance of production system.

3.4 Data Manager

The data manager provides a coordinating and oversight function over the data resources of multiple applications. The data manager may also, at the request of the project manager, provide direct support to a specific project. This support could include: requirements development, data modeling, planning, metadata development, and quality assurance. See "Role of the Data Management Team" in this document for further information.

3.5 Data Administrator

Data Administrators are a group of people appointed within the Information Technology area who jointly plan and administer the format, distribution, and storage of data. In this capacity, the Data Administrator translates the requirements of the Data Steward into a technical specification which can be programmed into an application system. The Data Administrator typically:

a) creates models of the system showing the processes and data stores it will contain;
b) creates the logical organization of data elements and related metadata which forms the basis for physical file layouts and database schemas;
c) provides technical reviews and testing to ensure that the software correctly processes the data; with the Data Steward, establishes the plan for migration of legacy data to the new system;
d) manages the technical aspects of the protection and delivery of production data; and,
e) maintains the model and technical documentation during the deployment and maintenance phases of the system.

Data administration for the Service Centers is typically a partnership between the software developers, the data stewards (the persons responsible for the content, accuracy, and completeness of the physical data) and the Data Administrator (the person responsible for planning and managing the physical storage, access, and descriptive information (metadata) for data).

As currently organized, the individual administrators also act within a broader managerial body referred to as the Service Center Data Team.

If the data administration function is carried out fully:

a) users have ready access to business information
b) procedures are in place to share information while maintaining data versioning and stewardship
c) a unique identity for datasets and data elements is maintained
d) the mechanics of accessing data is made as transparent as possible for the user

    e) an environment which facilitates software development and software implementation is maintained

3.6 Database Administrator

The Database Administrator is assigned overall responsibility and authority for the operation of a database. Their major responsibilities include:

    a) act as the designated authority for decisions concerning the modeling, construction and operation of a database
    b) coordinate with system developers to establish the most effective database software, hardware configuration, and data storage distribution.
    c) support the data steward in implementing customer access to data.
    d) develop technical procedures and components for storing and accessing data and metadata.

3.7 Application Developers

The analysts, designers, modelers, and programmers for an application system will:

    a) Model the data used and stored by the application.
    b) Provide application-specific metadata to describe data generated and stored by the system.
    c) Ensure that existing data is used in the application wherever appropriate.
    d) Engineer the data created by the system to be sharable by other applications as appropriate.
    e) Coordinate with database administrators to ensure physical database designs are efficient and tuned for the specific database and environment where the application will run.
    f) Prepare the data management deliverables specified in the Project Data Management Plan
    g) Implement data security and privacy rules within the application.

3.8 Data Collector

The Data Collector is an agency staff member, or a contractor, who is assigned the task of gathering data for a particular geographic area, for the area served by an office, or for a particular business process (such as processing loan applications). The data collector will:

    a) Gather data to meet the business area mission.
    b) Collect and document metadata required to describe data being collected.
    c) Ensure the accuracy of data and metadata to the extent possible.
    d) Ensure accuracy in the initial entry of raw data into automated systems and databases.

3.9 Data User

This category consists of all persons that use the data assets, including service center staff, service center customers, partner organizations, state and local governments, outside users of agency information, members of the agency business areas, and IT management and staff. The data user will:

    a) Use data in the way it is intended to be used.
    b) Take responsibility for finding out the proper definition and usage of data.

    c) Provide information that allows data related to the user to be extracted and correlated.
    d) Take steps (security, login ID's, etc.) necessary to establish access to data stores.
    e) Provide thoughtful feedback to application developers and data stewards on the quality, utility, and timeliness of data.

3.10 Service Center Modernization Initiative Data Management Team

The Data Team is an interagency team established by the Service Center Agencies. The team is responsible for implementing data management principles, policies, standards, and for establishing the overall data architecture.

The Data Team provides:
  a) a core data architecture, to include
    - Maintaining the Enterprise Data Model for all new/reengineered applications.
    - Coordinating the collection of metadata for spatial and tabular data.
    - Maintaining the business rules supporting the Enterprise Data Model.
    - Modeling the physical layout and location of national data used by the Agencies.

  b) strategic planning for the acquisition and use of data assets to meet program goals.

  c) data administration functions for applications, to include:
    - Resolution of conflicting data names, establishing common lookup tables, setting common domains for sharable data elements, and establishing unique keys and identifiers.
    - Coordinating data administration/management training.
    - Coordinating the release of data to the public.

  d) management of the ongoing data management process, to include:
    - Reviewing, at a minimum annually, and recommending modifications to the data management policies, as appropriate.
    - Developing standards, procedures, and shared utilities and tools for data management.
    - Maintaining a shared Data Management Handbook that documents common data management standards and procedures.
    - Coordinating implementation of a metadata repository, CASE and modeling tools, and other supporting data management software.

  e) a shared, central metadata repository for use by the Agencies to store and provide access to metadata, to include
    - Establishing standard and sharable data elements to promote data reuse.
    - Making metadata accessible to system developers and other users.
    - Maintaining a central source of data name abbreviations, and common acronyms.

  f) a consolidated voice to the department and other government committees on data management issues.

  g) Coordinate and support database functions, to include:
    - Coordination and review of logical and physical models.
    - Providing performance measures and metrics.
    - Implementing and coordinating security rules.
    - Monitoring databases for performance.
    - Coordinating and supporting reconstruction of databases to accommodate new information or to facilitate changes in the physical deployment of the database.

  h) Coordinate and support establishment of data warehouses to include
    - Coordinating data that crosses business areas.
    - Coordinating the development and maintenance of the warehouse data model.

4.0   Standard Data Element and Common Tables

Consistent usage and definition of data elements enables data sharing and integration of applications across the partner agencies. To this end, the USDA Service Center partner agencies maintain a set of standard data elements and common lookup tables. Data elements are added to this standard set when a single, concise definition can be determined. In cases where a given piece of data can have multiple nuances in meaning, multiple standard data elements with different names are created, each with one flavor of definition. In a similar manner, domains of values that further define a given data element are maintained. These domains can be used in lookup tables to validate data input, or can be used to populate choice lists on data entry screens.

The standard data elements and lookup tables are housed in a central metadata repository managed by the Service Center Data Team, and are available for use by all software developers.

4.1 Charge to Software Developers

   a) All software projects will use existing standard data elements whenever possible. A new data element will not be created if an existing standard data element with the same definition already exists. All new data elements will be developed with the idea of making them shareable with other new applications.

   b) All projects will use the existing standard lookup tables wherever possible. Because of the high cost of maintaining tables of values, a new lookup table will not be created for a project if a table already exists. All new lookup tables will be developed with the idea of making them shareable with other new applications.

4.2 Procedure

1. Upon determining the requirements for new system, the developer will consult the central metadata repository to determine a) what data may already exist, b) what business group should be contacted regarding use of the existing data.

2. The developer will obtain from the element/table owner a copy of the metadata and business rules for the sharable data elements or tables, and ensures that the new system will not alter the meaning and validity of the data/table.

3. While data may be sharable, there are still rules of "ownership" that must be established and abided by. Any system that writes to a data element must take into account who is dependent on the data, the versioning or temporal requirements for the data, and make a commitment to maintaining the physical data for as long as systems are dependent on it.

4.3 Reference

See the "Software Developer's Guide to Standard Data Elements and Common Lookup Tables."

5.0   Glossary

Business Area

An authorized program function or mission within an agency for which managerial responsibility has been delegated to an individual.

Business Rule

A statement that defines or constrains some aspect of the business as it is implemented in the data model (e.g. "an agency office can exist in only one location at a time").   Data-related business rule s are statements, phrased in absolute terms, about data (e.g. "a telephone number must have 10 digits"), and about relationships between data (e.g. "if a phone number is entered, the phone type must also be entered").

Change Control

Change control is an active management process stabilizes the software development environment and protects systems/users who are dependent on particular data from being adversely impacted by changes to the definition, type, availability, or content of electronic data.   The basic prerequisites for providing change control are: 1) establishment of management control over the definition of the data; 2) a means to notify users that a change is being proposed; 3) a means to gather responses from dependent users and to assess the potential business impacts of the change.   Note that the users may be programmers on the project who are dependent on the format, availability, metadata, and stability of the data stores to complete their tasks.

Clearinghouse Metadata

Metadata that describes a physical instance of data that describes another set of data. Clearinghouse metadata is used to describe and locate physical instances of data or software. Clearinghouse metadata can describe relational databases, data files, documents, photographs, video segments, stored sounds, software systems, applications, software components, libraries, etc. The level of granularity of the metadata may vary. In one case, the metadata may describe actual data objects. In other cases, the metadata may contain pointers to other sources or repositories of metadata. One set of metadata may list all available photographs and where to find them. The metadata describing each photograph in detail may be stored elsewhere with the actual photograph.

Common Data

Data jointly owned, used, and managed by Service Center partners. All partner agencies can write to and update the dataset.

Component

A reusable software module that encapsulates a set of behavior, hiding implementation details.

Component-Based Development

Creating applications by assembling existing components.

Data

A discrete fact or value. Data is the raw material, which through its use and interpretation can provide valuable information. Data is the content of databases or data files.

Data Administration

Data Administration encompasses the day-to-day technical functions that support ongoing business operations. Each application must implement a data administration process to support system

development and ongoing system operation. It includes the collecting, defining, certifying, organizing, protecting, and delivery of both data and metadata (data about data).

Data Administrator

The person who defines, organizes, manages, controls, protects and standardizes data models, data elements, and metadata.

Data Architecture

An orderly arrangement of data resources to achieve: 1) a common understanding of available data resources; 2) a planned approach to data acquisition, storage and retrieval to achieve a high degree of responsiveness to user demands; and 3) a high degree of data sharing and data mobility to reduce program delivery costs.

Data Clearinghouse

A facility for advertising and distributing datasets. Metadata describing available datasets is made available to a chosen audience (i.e., the general public, partner agencies, etc.). Packaged datasets (i.e., diskettes, CDs, etc.) or online retrieval of selected subsets of data are distributed within security guidelines. There may be a charge for obtaining datasets.

Data Dictionary

A database about data and database structures. A catalog of all data elements, containing their names, structures, and information about their usage. Normally, data dictionaries are designed to store a limited set of available metadata, concentrating on the information relating to the data elements in the databases, files and programs of implemented systems.

Data Integrity

The state that exists when data is handled as intended and is not exposed to accidental or malicious modification, destruction, or disclosure. Data is entered using formal business rules, and accuracy is checked using automated edits and manual reviews. Also, includes the preservation of data for its intended use.

Data Management

Data Management is the managerial function of taking responsibility for data and the processes that support it. It focuses the strategic planning and data methodologies for meeting program delivery goals. In particular, Data Management aims at managing data as an asset, particularly as a corporate asset. Data managers typically look across applications and business areas to manage the whole architecture of data resources for the enterprise.

Data Mart

A type of data warehouse that contains smaller subsets of data and focuses on a particular business discipline or organizational component.

Data Model

A pictorial view of data, groupings of data, and relationships between data groupings. A "logical" data model is a view that does not depend on the characteristics of the computerized system or of the physical storage. A "physical" data model typically refines the logical model by adding the constraints incumbent to the database system or physical storage method, and tuning the data model for access efficiency. A "business" data model typically identifies the main categories of data used by and created by the application, integration points with other systems, data sources external to the enterprise, and known data structures that will be shared. A

15

"conceptual" model is a further refinement of the "business" model, with greater detail, but not detailed specifics on data elements, tables, and other data stores.

Data Repository

A database of information describing the characteristics (metadata) of data. Typically, the repository also stores a broad range of descriptive information, including business rules, data models, and process models that help to elaborate on the usage of data in various systems. Repositories can also store metadata for the purpose of identifying and retrieving sets of actual data. Metadata that describes a map is an example.

Data Steward

A business area expert who is assigned responsibility for the data content of the database. The data steward establishes business rules, defines data elements, identifies valid data values, establishes certification standards, and ensures the completeness and availability of the data.

Data Validation

Applying a set of rules, comparisons, or decisions to a data element to determine if it falls within the pre-established boundaries of values for that element.

Data Warehouse

An informational database, or collection of databases, used to store shareable data. The warehouse is usually created through data extracts from operational databases. The warehouse adheres to a single enterprise data model to ensure consistency of decision-support data across the enterprise. The warehouse typically allows users to tap into an organization's vast store of operational data to track and respond to business trends, and to facilitate forecasting and planning efforts.

Data Warehouse Metadata

Metadata that describes the contents of a warehouse. This metadata can include the original source of the data, timestamps, data conversion routines, data transformations, volatility of the data, refresh periods, data reliability indicators, relationships between data from multiple sources, etc. Warehouse metadata can also include metadata describing individual data elements stored in a warehouse.

Database

A collection of related data organized to serve one or more applications. In the broader sense, it describes any organized collection of data regardless of the physical storage method.

Database Administration

The function of designing, implementing, securing, operating, and maintaining a collection of data in a database management system (DBMS). This includes the implementation of rules by which data is accessed and stored, performance monitoring, data backup and restoration, and management of the DBMS software and its environment.

Database Administrator

The person who creates, manages, controls, and protects a database.

Domain

A list of all possible valid values for a data element. The domain can alternatively be expressed as a range of numeric or alphabetic values, or as a reference to an identified standard, such as a FIPS table.

Enterprise Data Model

An overall pictorial view of the participating agency's many applications and data assets. The intent is to manage the overall data assets to achieve optimal integration, sharing, access, and utilization of technology resources and infrastructure.

Geospatial Data

Information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. This information may be derived from sources such as remote sensing, mapping and surveying technologies. It includes both attribute (text) as well as spatial (map) data.

Geospatial Metadata

Metadata that describes geospatial features. Points, lines, and polygons are created in a GIS tool and have little meaning if they are not described in detail. These descriptions can include such items as a name for the feature, category (i.e. farm, field, water, wetland), creation date, crop cover, ownership, bounding coordinates, scale, etc. There is geospatial metadata, as well, that describes large sets of geospatial data. This includes metadata that describes a complete map. This type of geospatial metadata is used in particular to advertise or allow users to locate maps and geospatial datasets.

Information

A commodity derived from data through analysis or by the orderly presentation of data for human interpretation.

Lifecycle Phases

Initial Proposal - A description of the business objectives to be attained, and what impacts the project will have.   The proposal establishes the system's intended place in the enterprise architecture, identifies resources that will be required, establishes essential roles in the Information Technology and Business communities, and provides essential information to support management decisions regarding the value of the proposed system and the level of resources to commit.

Requirements - A complete description of the characteristics, capabilities, capacities, and information a software system must contain to meet the needs of the business and the end user.

Design - A plan for the characteristics and construction of items of hardware, software, data, and procedures to deliver the business processes described in the system requirements.

Development - The process of turning a system design into a working system.

Testing - A period of rigorous exercising of system components to determine the accuracy of system processes and the capability of the system to coexist and function adequately in a production environment.

Deployment - The process of readying a system and its data stores for production operation. Making the system available for use by the end-user community.

Maintenance - Correcting processing errors discovered after deployment, and incorporating minor changes in hardware and software needed to keep pace with changes in technology or availability of commercial products and resources.

Local Database

A database, which is used by an agency subunit (e.g. Region, State, Center, Institute, etc.) but does not meet the definition of a National Database. Examples include: local spreadsheets, report extracts, and PC-based databases.

Metadata

Data about data. Metadata describes how, when, and by whom a particular set of data was collected, and how the data is formatted. Metadata includes attributes such as data name, length, domain of valid values, and definition. Metadata can also identify and describe a set of data or a complex data type such as a map, photograph, spatial data set, etc.

Metadata Repository

A database of information describing the characteristics (metadata) of data. Typically, the repository also stores a broad range of descriptive information, including business rules and data models that help to elaborate on the usage of data in various systems. Repositories can also store metadata for the purpose of identifying and retrieving sets of actual data. Metadata that describes a map is an example.

National Database

A permanent database that (1) has international, national, USDA, or agency-wide application, (2) is included in a standard software suite, (3) contains data that is used/shared directly in making national program decisions, or (4) is used/shared in multiple offices, states, or other internal/external organizations.

Registry Metadata

A general term for metadata that gives data elements a name and definition, describes how the data will be stored, records relationships between data items, captures the business rules for data element and entities, describes the domain of valid data values, and identifies the caretakers of the data. It may include data models and database designs for specific application(s) where the data is used.

Shared Data

Data owned and managed by a specific Service Center partner and shared by other partners. Usually, only the owning agency writes to the dataset, while other agencies can read the data.

Standard Data Element

An element or structure that has a definition acknowledged by all partner agencies.

Strategic Data Management Planning

Planning how data is acquired, stored, and used in the most efficient manner to support the agency mission.

Unique Data

Data owned and managed by a specific Service Center partner and not shared. The data is usually mission-specific, and is written and read only by the mission area of the particular agency.

Appendix A – The Software Development Life Cycle

There are many varieties of life-cycle methodologies for application development projects. In this document, a generic software lifecycle is utilized which encompasses the general phases found in most projects. The phases are Initial Project Proposal, System Requirements Development, Physical Design of the system, Development and Coding, Software Testing, and the Deployment and Maintenance of the System. The Maintenance Phase can be quite lengthy and involve numerous changes and upgrades to the basic system over time.

There may be multiple iterations of the of the development lifecycle within the lifetime of an application.
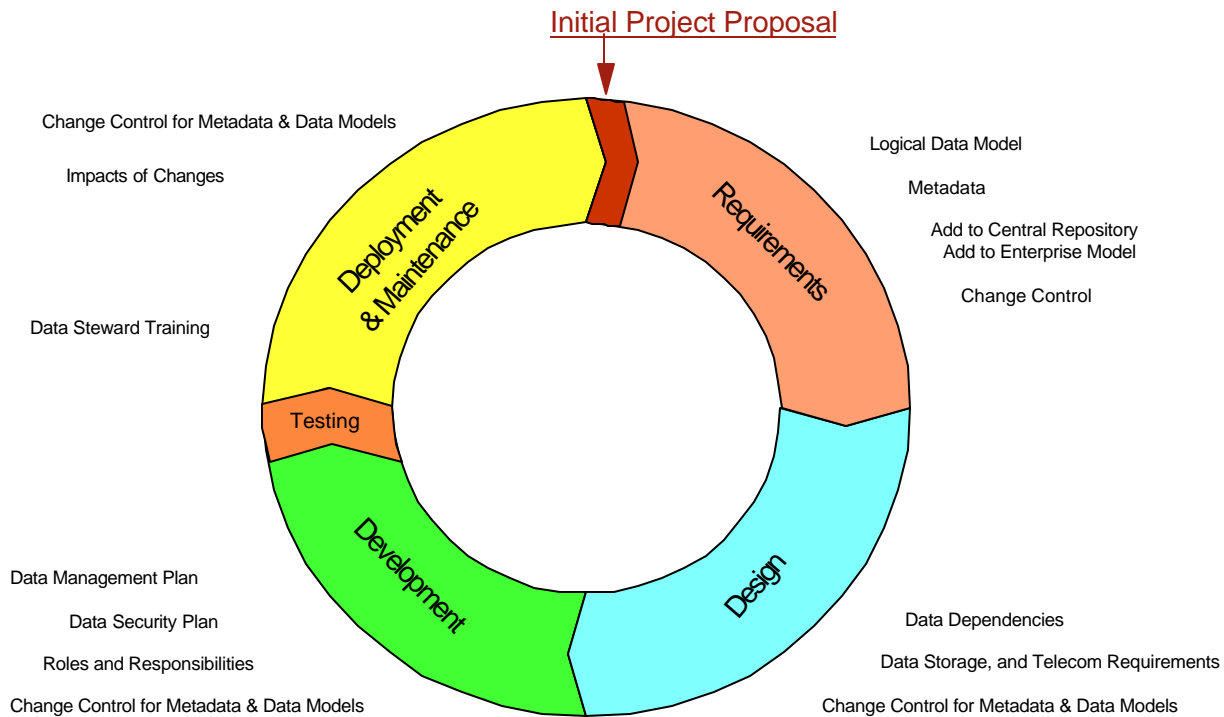

Exhibit 1 – Data Management Deliverables in each Lifecycle Phase

Appendix B - Data Management Deliverables for Application Development Projects

Note: This section is drawn from the document "Service center Data Management Deliverables for Application Development Projects."

The following pages list the project deliverables for each phase of the application development lifecycle. The deliverables either require a written document or an activity such as uploading of information.   These requirements have been pared down to a bare minimum needed for overall enterprise management of data resources. It is the responsibility of the Service Center Data Team to provide this overall enterprise management and to ensure coordination across the large slate of information system projects. The Data Team will be provided copies of each deliverable for review. However, each deliverable is also an essential part of system documentation, and will also be maintained with the project's documentation set.

Preparation of these deliverables is seen as a joint responsibility of the Application Developers, the Data Steward representing the sponsoring business-area, and the Service Center Data Team. There may be other data-related products needed by the application developers and data stewards to support a system in addition to those detailed in this document.

A checklist of additional data management activities that should be accomplished during each phase of project development is included.

Data Management Deliverables for the:

Initial Project Proposal

The following chart lists Data Management deliverables to be provided in the Initial Project Proposal or for a Waiver Request by an application project team. This Proposal establishes the system's intended place in the enterprise architecture, establishes essential roles in the Information Technology and Business communities, and lays out the basic impacts of the system. Whether the system will be implemented using commercially-available packages or through a full software development project may not be known when this proposal is submitted.

	D E L I V E R A B L E S
1.	**Establish project roles**, including who in the business area will be the Executive Sponsor of the system, the Project Manager, and the Steward of the physical data and metadata. Include the person's position title. The Data Steward is a person from the business area.
2.	**Identify special data types** that the application will potentially require. These are complex data types such as photographs, scanned documents, digital signatures, spatial data, etc. that may require the purchase of new database, network (including browsers), or client/server software; or additional storage capacity.
3.	**Describe special data security and data handling requirements** for the system.
4.	**Identify data dependencies**. Will data from an existing system continue to be used or brought forward to the new system? Will other systems be dependent on data from the new system? Will the new system be dependent on data from other systems? Developers must make a good-faith effort to locate existing electronic data stores before requiring the manual origination of data.
5.	**Create a high-level Business Diagram** with supporting documentation indicating the main categories of data used by and created by the application, integration points with other systems, data sources external to the enterprise, and known data structures that will be shared. Indicate the types of data, such as tabular, spatial, document, binary, etc.

Data Management Review Checklist:

Y   Add a placeholder in the Enterprise Model for the application. Begin to identify existing sources of data to meet requirements outlined in the 'High-level Business Model' and the 'Data Dependencies'.

Y   Coordinate requirements for special data types and special security needs with other projects that may have similar needs.

Y   Discuss initial ideas about the technical environment envisioned for the system. The focus is on early identification of any large or highly dispersed data storage requirements so as to forecast potential load on telecommunication networks and data storage equipment; and to begin any necessary procurement processes.

Y   Identify data management resources and technical knowledge that are needed but will likely be unavailable to the project, particularly to complete the requirements and design phases. This can include skills such as data modeling, database design and tuning, warehouse design, data management, etc.

Data Management Deliverables for the:

Requirements Phase

The requirements phase is where detailed business needs are identified and where system analysts convert these needs into technical system requirements. Data stores to support the application are identified, along with the individual data elements they will contain.

	D E L I V E R A B L E S
1.	Develop a **logical data model**. Use a CASE tool to graphically portray data elements, how they are grouped and existing relationships between elements or groups of elements. Models are not confined solely to relational databases. A flat file is a grouping of data elements, and there can be relationships between files. Hierarchical databases can also be modeled. The logical model will be incorporated into the Enterprise Data Model.
2.	Create **metadata** to describe data elements, database tables, files, geospatial features, business areas, systems, and subsystems. The required contents for this metadata is described in the *Service Center Tabular Metadata Standard* and the *Service Center Geospatial Feature Metadata Standard*. This metadata is based on the logical data model, and will usually be captured by the project team in a CASE tool. Note that metadata for data elements is created by a project team only for new elements. If data is being used from another application, that application will have already established the required metadata. There are also established Standard Data Elements which should always be used before creating new application-specific elements. These Standard Elements can be obtained from a Data Team representative. The names of data elements will conform to the *Service Center Data Naming Standard.* For COTS-based applications: When the content of the internal data stores supporting a COTS package can be accessed by a project developer, elements in that database will be included in the dictionary. Data element and data element group names established by the COTS package are not required to follow USDA naming standards.
3.	Upload the **Metadata** and **Logical Data Model** to the central metadata repository.
4.	Establish **Change Control** over all metadata and the logical data model.

Data Management Review Checklist:

Y   Application Developers will review the Logical Data Model with a Data Team Representative. This is best done periodically throughout the requirements phase, rather than after the model has been "finalized".

Y   Actively search for an existing electronic source for each data item before committing to collect the data.

Y   Incorporate the project's logical model into the Enterprise Model.

Y   Application Developers will review all metadata with the Business-Area Data Steward and the Business-Area Data Administrator.

Y   Upload metadata describing data elements, data element groups, subsystems, systems, geospatial features, business areas, and stewardship to the central metadata repository. The uploaded metadata will likely be replaced by more complete information in later project phases.

Y   Identify roles and procedures for maintaining change control over metadata and the data model.

Data Management Deliverables for the:

Design Phase

The system requirements have been fully described in logical terms, and must now be cast in physical terms to send to the programmers. Now that the system is visualized in detail, it is important to reevaluate its relationships with the outside world.

		D E L I V E R A B L E S
1.		**Formalize Data Dependencies** with other systems. Establish and maintain a contact list of all application systems and end-users who will have a stake in the data output from this system so that users can be notified before any modifications to the data definition, domain, or content are made. Data elements being shared from/with other systems are formally declared. Agreements should be formalized with providers of any shared data resources on which the new application is dependent.
2.		**Maintain Change Control** on all data items in the central metadata repository, data model, and enterprise model. Ensure that all changes to data items made during the design phase are recorded in the metadata, and that dependent users are notified in advance of any changes.
3.		Provide **Detailed Requirements** for data storage, telecommunication loads, and software/databases needed to support special data types. Projects should look at their expected geographical dispersion of data, data volumes, and processing loads, including the estimated average and maximum data set size and transmission frequency.

Data Management Review Checklist:

Y   Map data dependencies between systems into the Enterprise Model.

Y   Establish data structures and keys to allow for efficient movement of data between computers, for the combining of data if two offices are merged, to allow data to be extracted as a complete set, and for downloading of an office-worth of data from a national or regional server.

Y   Consider georeferencing of data for mapping and graphical representations of management information.

Y   Analyze expected data loads in conjunction with telecommunications and equipment management teams.

Y   Maintain integrity of information in the Central Metadata Repository and the Enterprise Model.

Y   Make metadata, including a data dictionary, readily available for use by developers in the next phases of the project.

Y   Employ specialists to denormalize and fine-tune the physical data model for system efficiency.

Y   Establish the complete data environment, including model and definitions, before beginning the coding phase.

Data Management Deliverables for the:

Development Phase

This phase concentrates on: 1) documenting the system as programmed; and 2) keeping the documentation up-to-date as technical refinements are made. The Detailed Data Management Plan and the Security Plan are normal documentation produced for any project.

	D E L I V E R A B L E S
1.	Maintain **Change Control** on all data items in the central metadata repository, data model, and enterprise model. Ensure that all changes to data items made during the coding phase are coordinated with dependent users, and recorded in the metadata.
2.	Develop a detailed **Data Management Plan**. This is a detailed plan for the ongoing management of data within the business area and the application. This plan communicates roles and responsibilities to users and maintainers of data in the operational system. It sets the rules for certifying the accuracy of collected data, and for protecting data wherever it is stored. Sections of this plan may be implemented through the application's user guides and other end-user documentation. The Business-area Data Steward has primary responsibility for this deliverable.
3.	Formulate a **Data Security Plan**. Analyze the risk involved in the loss or corruption of data. Describe the data produced by the system, its lifespan, data retirement procedures, backup and archiving, physical data security requirements, and access rules. Include data integrity rules for preventing accidental loss of data, data corruption and the inappropriate updating or modification of data as it is transferred between offices, computers, and systems. Describe requirements for data mirroring, and for disaster recovery. Describe how record and spatial data key clashes will be handled. Describe how concurrent updates will be handled.
4.	**Establish detailed Roles and Responsibilities** to be performed by the system's Data Steward to maintain both the physical data and the metadata for the application. There may be additional Data Stewards responsible for portions of the data, such as a State or Local Data Steward. Also describe any additional roles to be performed, such as Database Administrator, Security Officer, Data Collector, Data Certifier, Steering Body, State IT Staff, etc.

Data Management Review Checklist:

Y   Maintain lines of communication between developers and the Service Center Data Team to ensure the integrity of the metadata and models.

Y   Perform a joint review of the Data Management Plan, the Data Security Plan, and the Roles & Responsibilities.

Data Management Deliverables for the:

Deployment and Maintenance Phase

This phase concentrates on: 1) documenting the system as programmed; and 2) keeping the documentation up-to-date as bug-fixes and system enhancements are made.

	D E L I V E R A B L E S
1.	**Maintain Change Control** on all data items in the data dictionary, enterprise model and central data repository. Ensure that all changes to data items made during the deployment and maintenance phase are recorded. Maintain all data-related changes in the project CASE tool and in the central repository.

Data Management Review Checklist:

Y   Coordinate the training of Data Stewards to emphasize their role in protecting the integrity of data stores wherever they are located in the country, and for ensuring the validity of data input into the system.

Y   Fine-tune and update the Data Management Plan for the system.

Appendix C - The Data Management Team

C.1   Roles

The Data Management Team, as a shared coordinating body among the partner agencies, has a number of roles:

1. <u>Standard Setter</u> - use the collective knowledge of data management professionals in multiple agencies and business areas to set basic standards for the handling and documenting of data.

2. <u>Coordinator</u> - use the approach that consensus, while taking longer, often builds a more lasting agreement. Data management is important, but we are still in the beginning of the window of opportunity. Movement, wider-spread usage, and reuse of data provide the real return on the investment in data management. Systems and business philosophy are just beginning to incorporate these concepts. The work already done by the Data Team puts them ahead of the curve. The big opportunity is to be ready when systems come full-circle in their lifecycles and are ready to reengineer. Continued coordination activities with developers and between agencies will highlight opportunities for data management wins.

3. <u>Architect</u> - Seeing the bigger picture and translating it into an overall architecture in which systems and data stores can be built with minimum cost and minimum duplication of effort.

4. <u>Service Provider</u> - the long-term win is not in telling developers to do something, but in providing a service which the IT organization finds valuable and essential. If the Data Team provides direct hands-on services which add value or reduce development time, the policies and procedures for good data management will be better accepted into the mainstream development lifecycle. "Lead by example." A good programmer often has poor data skills. Skilled data administrators can provide high-quality modeling, normalization, metadata management, and database management support.

5. <u>Enforcer</u> - So far, not a big crowd pleaser. System developers adopt easily what adds value. Inroads have been made for tools that work within individual projects. the challenge is to show how the activities within individual pr ojects provide sufficient advantages to the larger environment, i.e. how standardization and change control add value to the enterprise environment and who reaps the benefits.

6. <u>Procurer</u> - Doing the market research and procurement processes for data management tools and databases. Procured contractor support for various initiatives.

7. <u>Researcher</u> - Doing the research to discover good data management practices being implemented in corporate settings and other government agencies. (databases, data warehousing, data query tools, XML, etc.)

C.2   Data Management Benefits to the Agencies:

1. Better data management provides better program delivery. With greater use of COTS software and networks, more of the work for keeping systems operating efficiently will fall on the data side.

2. Data collection, validation, and certification is usually the most expensive aspect of information systems. To the extent that efficiency is improved, operating costs are reduced.

3. As the WEB is proving, data is becoming the product rather than the by-product. Its care and packaging is making increased demands on business staff resources.

4. For those systems that move out of the office and onto the farm, the fluid movement of data will make or break the utility and efficiency of the effort. If the process is cumbersome, the advantage is lost. Careful data management planning is essential.

5. With reduced resources and improving electronic communication, more of the responsibility for participation in USDA programs will fall on the customer. Good data management practices are essential to e-government.

C.3   What the Data Management Team can do in the Future

We have the opportunity to focus on a few areas which have a potential for high return. Many information initiatives have come and gone. Yet the one lasting thread across all of these initiatives has been the infrastructure put in place to foster and support sound, progressive data management.

The following have the potential for a good return on the investment in staff time:

1. Support the implementation of SCIMS, OIP, CAMS, and the Geospatial Themes as an integrated foundation on which to build future BPR initiatives.

2. Become proactive in the implementation of E-government.

3. Continue the development of standards.

4. Take a proactive roll in the development of data warehousing, particularly as a means for delivering data to a variety of customers.

5. Maintain the data team website and repository, and promote their usage.

6. Actively promote the use of standard data elements for application development. The long-term goal from a data enterprise perspective is to make data more shareable and better organized for use in the future "connected" society.

7. Provide direct support to projects to ensure good data management techniques are incorporated from the very beginning.

C.4   Specific Data Team Tasks

Data Administration Support
- Enforce Standards
- Provide modeling and metadata support to application developers
- Support business areas of the agencies in establishing data requirements, and implementing data administration roles and responsibilities within their organizations.
- Monitor data stewardship functions throughout the agencies.
- Provide administrative support for change control of data elements.
- Provide cross-divisional coordination of data administration for ITS.

Tools and Training to support Data Administration functions
- Provide user training on CASE tools; Data Administration principles and standards; and roles & responsibilities for developers and business users.
- Refresh CASE tool technology as needed.
- Maintain a central metadata repository with interfaces to CASE tools.
- Establish shared metadata search engines and WEB links, as appropriate.

Data Sharing and Access
- Continue building and maintaining the Enterprise Data Model.
- Establish frequent reviews of application development deliverables to ensure data is shared and reused as appropriate. (Waiver process support)
- Support initiatives to enhance user access to data.
- Support data security.

Planning
- Evaluate and implement emerging technologies for administering data.
- Establish long-term plans for integrating applications; implementing effective data storage; maintaining data security; refreshing policies and standards; and interfacing applications with data warehouses.
- Establish methods to improve data administration processes

Stewardship of Common Datasets
- Establish and support datasets shared across divisions and business-areas.

Database and Warehouse Support
- Provide logical modeling services to support the implementation of physical databases and data warehouses.
- Establish a logical Enterprise model for warehouse initiatives across the organization.

Appendix D - National and Local Data Stewards

Data Stewardship occurs wherever data is stored. Each application will have a national steward who reports directly to the project's Executive Sponsor. The national steward is involved in the initial creation of the system, as well as the ongoing operation of the implemented system. If data is collected and stored at state and field locations, local data stewards will be appointed for those portions of the data.

D.1   National Business-Area Data Steward

The Data Steward is responsible for the content of a database or dataset. The Data Steward establishes definitions and domains for data elements; sets the procedures for collecting and certifying data and metadata; and manages the overall storage, maintenance, and distribution of the data and metadata. This person often acts as a conduit between the end-user community and the IT community. The Data Steward may have frequent contacts with data administrators, application development teams, and database administrators supporting the business area.

The Responsibilities of Business-Area Data Steward:

1. Manage/supervise/administer data needed to support the mission.

2. Authorize individual customer access to data and set any limitations on that access (read/update/access to subsets/etc.).

3. Determine which reports/data-extracts represent 'official' agency information.

4. Resolve disputes as to the meaning and valid use of data elements and values.

5. Cooperate with fellow Data Stewards and the Data Team to develop/modify/approve/enforce policies relating to the use and distribution of data resources.

6. Formally delegate Data Steward's duties to others where appropriate so that individual responsibilities are clearly defined.

7. Act as the designated authority for business-area decisions concerning data content and requirements for supporting software systems.

8. Develop technical procedures for acquiring, collecting, archiving, and disseminating both data and metadata, including establishing cooperative agreements with outside sources of data.

9. Develop and implement defensible quality assurance standards for data collection and management, and a quality control process to certify that standards have been met.

10. For geospatial data, ensure that metadata is documented, approved, certified for release and made available through the Clearinghouse network according to the current FGDC metadata standards.

11. Coordinate with the Security Officer to identify security requirements for any information that may be excluded by the provisions of the Freedom of Information Act or must be protected under the Privacy Act.

12. Identify training needs, develop training plans, develop training materials, and train users on use and management of the data.

13. Certify that software applications meet subject-matter and technical accuracy requirements.

14. Coordinate with other USDA Agencies and Offices in setting development priorities and activities, and data accuracy and usage standards.

15. Provide "help desk" support to governmental and outside users of data resources.

C.2   Local Data Steward

A local Data Steward will be assigned at each point in the organization where data is stored. The same person can be assigned for multiple storage points, or it can be different people. Data cannot just sit around without someone being responsible for its use, security, and integrity. If data is widely distributed, stewardship must expand to cover it. Stewardship responsibility follows the data wherever it goes.

'Local Data Stewards' are delegated specific responsibilities by the Business-Area Data Steward. These responsibilities may range from control and oversight of data collection and maintenance for an entire State to protecting a copy of a small subset of data used within a particular Service Center. Local Data Stewards should come from the business-area staffs at regional, state, area, or local offices.   However, this certain responsibility may be delegated to someone outside the business-area, including an Information Technology person or a data user.   At local levels, for example, this responsibility may be assigned to a Service Center employee who is not directly part of the business-area, but who uses data from multiple business areas in providing services to the public. All delegations of responsibility for data should be done formally, either by letter or through position descriptions and performance elements.

When copies or subsets of data are stored at state, regional, or county offices, at universities, or with local governments, the Business-Area Data Steward and the Executive Sponsor need to establish 'Local Data Stewards' to maintain these datasets. Data assets must always be within the span of control of the responsible business area, or their delegates, wherever this data is located.   Copies of data may be released to outside parties, but only under conditions established by the business area to ensure the integrity of the data and the protection of privacy, data sensitivity, and security.

Collaboration and communication must take place among data stewards at all levels of the organization to ensure that data assets of the enterprise are protected, maintained, and used as intended.   This is particularly true when data moves between machines and offices.